AD_____

Award Number: DAMD17-01-1-0376

TITLE: Investigating the Mechanisms of Action and the Identification of Breast Carcinogens by Computational Analysis of Female Rodent Carcinogens

PRINCIPAL INVESTIGATOR: Albert R. Cunningham, Ph.D.

CONTRACTING ORGANIZATION: Louisiana State University
Baton Rouge, LA 70803-2701

REPORT DATE: August 2003

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 2003	Annual (15 Jul 2002 - 14 Jul 2003)

4. TITLE AND SUBTITLE
Investigating the Mechanism of Action and the Identification of Breast Carcinogens by Computational Analysis of Female Rodent Carcinogens

5. FUNDING NUMBERS
DAMD17-01-1-0376

6. AUTHOR(S)
Albert R. Cunningham, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Louisiana State University
Baton Rouge, LA 70803-2701

E-Mail: acunningham@lsu.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

This project is investigating the potential that environmental estrogens may be involved in the etiology of breast cancer. We hypothesize that specific features of chemicals can be identified that are significantly associated with female and breast carcinogens and that these features are related to mechanisms of chemical carcinogenesis. Our overall scientific objective is to investigate the hypothesized relationship between environmental chemicals, xenoestrogens, and the development of breast cancer.

As described we moved from Pitt to LSU, causing a lag in progress during renegotiations. We are on schedule and budget and have accomplished the bulk of work for Specific Aim 1 (or year one). This includes MCASE SAR models for mouse, rat, female-specific, and mammary carcinogens. Due to multiple factors we are migrating our work from MCASE to Tripos Sybyl. Initial verification that Sybyl will perform the required tasks has been demonstrated.

We anticipate about two manuscripts will shortly be produced from Specific Aim 1, describing our female and mammary carcinogen model and the utility of Tripos Sybyl software to adequately deal with environmental toxicants.

14. SUBJECT TERMS
Structure-activity relationship (SAR), computer modeling, mechanisms and etiology of breast cancer, environmental carcinogens

15. NUMBER OF PAGES
9

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Annual Review August 2003

Investigating the Mechanism of Action and the Identification of Breast Carcinogens by Computational Analysis of Female Rodent Carcinogens

DAMD17-01-1-0376

Albert R. Cunningham, Ph.D.

Introduction

The well-established breast cancer risk factors may account for only 47% of the breast cancer incidence in the United States. This leaves a considerable portion of breast cancer from undetermined origin. This project is investigating the potential that environmental estrogens may be involved in the etiology of breast cancer. We hypothesize that specific features of chemicals can be identified that are significantly associated with female and breast carcinogens and that these features are related to mechanisms of chemical carcinogenesis. Our overall scientific objective is to investigate the hypothesized relationship between environmental chemicals, xenoestrogens, and the developmental of breast cancer.

Body

As reported in the first Annual Report, after working on this project for approximately six months at the University of Pittsburgh, I moved to the Department of Environmental Studies at Louisiana State University (LSU). The project has been successfully renegotiated at LSU. The budget has been redone with some extra time added as a no-cost extension to help make up for lost time. The project recommenced in the fall of 2002 at LSU. Therefore, although this is the second Annual Report, in actuality it basically covers the first whole year of work on the project.

Software Change

The structure-activity relationship (SAR) modeling was originally proposed to be conducted with the MCASE program. However, for multiple reasons, I have decided to switch platforms to Tripos Sybyl. This change does not alter the project and I am currently working with my grants manager, Dr. Moore, to update the SOW.

During the early part of this year it was becoming evident that MCASE was not developing models for this project (details discussed below) that were of stellar predictivity. On account of successful SIMCA modeling of aromatic amine *Salmonella* mutagens and skin sensitizing agents for a project supported by Proctor & Gamble, we spent some time investigating whether Sybyl could be employed to produce adequate models relating to this project.

Briefly, although Sybyl and MCASE are different modeling packages, the Sybyl family of SAR modules allows for a similar type of analysis of toxicants. As described in the proposal, MCASE takes a binary approach to analyzing toxicants by comparing structural features (2-dimensional biophores) found in active and inactive compounds. Similarly, for the Sybyl analyses the HQSAR (hologram QSAR) program calculates 2-dimensional holograms(i.e., linear fragments comparable to biophores) and the Advanced QSAR module uses the soft independent modeling of class analogy (SIMCA) algorithm to perform statistical analysis of the holograms. SIMCA is

a regression-type analysis that develops predictive models based on categorical data (i.e., carcinogens and noncarcinogens). Overall, the HQSAR-SIMCA models appear to be superior to MCASE models.

Our initial HQSAR-SIMCA included an analysis of potential carcinogens identified using ~1600 diverse *Salmonella* mutagens and 122 compounds tested for estrogenicity using the E-SCREEN assay for which MCASE models existed. The mutagenicity MCASE model and its analysis has been published (*1, 2*) and the E-SCREEN environmental estrogen model has been accepted for publication with minor revisions (*3*). In both instances, Sybyl has been able to develop models comparable to the predictivity of MCASE. We anticipate a manuscript from this work describing how HQSAR-SIMCA can be successfully employed for computational analysis and prediction of environmental toxicants.

At this juncture, all projects in my laboratory are being transferred to Sybyl. Needless to say, Tripos Sybyl software is, in my estimation, superior to MCASE. Notably, the Tripos family of software is growing and is on the cutting edge of technology. The literature is replete with Sybyl-based investigations. On the other hand, MCASE is owned by an individual and has the real potential to become a legacy or obsolete system in the coming years.

Specific Aim Accomplishments
The Specific Aims for year one are as follows:

Specific aim 1: Development and validation of SAR models for female breast carcinogens (months 1-12).

a. Identify chemicals tested in female rodents from the Carcinogenic Potency Database and the National Toxicology Program (month 1).
b. Enter chemical structures and potency values into MCASE program (months 2-8).
c. Validate models using 10-fold cross validation (months 9-12)
d. Summarize and interpret models and prepare publication.

These models have been developed and validated (i.e., a-c) as planned. Moreover, in conjunction with this, we have also taken the privilege to update our existing rodent carcinogenicity models so that all models (mouse and rat, as well and female specific version) have been built on the same datasets and analyzed with the same software version.

However, in summarizing and interpreting (i.e., d) it became evident that the models were not performing as well as anticipated (Table 1). Through a series of 10-fold cross-validations we calculated sensitivity (% carcinogens correctly predicted), specificity (% noncarcinogens accurately predicted) and the overall observed correct prediction (OCP) rate for each model. Looking at the shaded rows in Table 1, it is evident that the models developed generally had a low sensitivity and thus were not able to accurately predict carcinogens.

A common problem encountered with SAR model development and validation is that the model can only be validated against existing data. Additionally, data used in the learning set are rarely a complete and random sample of the universe of chemical features. Therefore, when using a SAR model to predict the activity of a novel chemical, uncertainty of the predictive ability of the

model exists when the chemical falls outside the sample space of the learning set. This problem manifests itself readily in MCASE during validation studies, particularly when the learning set is relatively small and multiple mechanisms are involved in the measured endpoint. Chemicals that contain a unique biophore (i.e., a feature represented only in the chemicals that are deleted from the learning set and placed in the validation set) are not accurately predicted since the model loses the chemical's unique informational contribution. These chemicals fall outside the sample space of the remaining model (i.e., outliers). This, in turn, lowers the sensitivity and

Table 1. MCASE model summary.

				MC Units			MC Percent		
Model	**Inactive**	**Marginal**	**Active**	**OCP**	**Sens**	**Spec**	**OCP**	**Sens**	**Spec**
CPDB									
Rodent	508	2	662	0.6456	0.6259	0.6713	0.6362	0.638	0.6339
Rat	442	0	473	0.6368	0.5508	0.7285	0.6171	0.5742	0.6629
Rat mod	442	0	372	0.7125	0.6532	0.7624	0.7014	0.6425	0.7511
Mouse	389	0	370	0.6245	0.5595	0.6864	0.6653	0.5216	0.8021
Mouse mod	389	0	279	0.7695	0.7384	0.7918	0.7695	0.7312	0.7912
[redacted]									
Rat Breast	87	0	88	0.7314	0.6932	0.7701	0.72	0.7045	0.7356
NTP									
Rodent	121	56	231	0.5784	0.5958	0.5372	0.5882	0.6202	0.5124
[redacted]									

Notes:
CPDB: models based on Carcinogenic Potency Database
NTP: Models based on the National Toxicology Program
OCP: observed correct predictions, number of correct predictions / total number of predictions
Sens: sensitivity, number of correct positive predictions / total number of positives
Spec: specificity, number of correct negative predictions / total number of negatives

concordance of the model (specificity remains the same). In order to assess this limitation in the validation procedure of MCASE and its derived model, all of the active chemicals that were identified by a unique biophore were removed from the overall validation set for a modified validation study. This procedure allows for the validation of a more robust model. These modified validation sets had an increase in sensitivity for rats from 55% to 65% and an increase for mice from 56% to 74% (Table 1).

Not surprisingly, the SAR model performance is enhanced by the removal of these single occurrence biophore chemicals. However, given the fact that some of the female specific models had sensitivities well below 50%, I began to question the ability of this type of outlier analysis to produce meaningful models of female-specific carcinogenesis. Therefore, models were generated using Sybyl.

Female Carcinogen Models

Specific Aim 1a is for the creation of female specific models. As discussed, these models have been developed for MCASE. We are currently transferring them to Sybyl for HQSAR-SIMCA analysis.

Mammary Carcinogen Models

Although rodent and female specific carcinogen models are important models for this project, the hallmark models are those developed for mammary carcinogens. Therefore, we thought it prudent to verify that Sybyl HQSAR-SIMCA would be capable of producing adequate models of breast carcinogens. We have established two HQSAR-SIMCA models for mammary carcinogenesis, one each for rat and mouse mammary carcinogens from the Carcinogenic Potency Database (CPDB) as part of Specific Aim 1. The mouse model, based on 48 compounds (50% mammary carcinogens and 50% noncarcinogens) is estimated to be approximately 81% predictive through cross-validation. The rat model, based on 200 compounds, is approximately 77% predictive.

These models are based on *all* mammary carcinogens in the CPDB including several male-only breast carcinogens. We will shortly be producing a model of female-only breast carcinogens. Also, as mentioned, the modeling technique compares carcinogens to noncarcinogens. In these mammary carcinogens models we compared mammary carcinogens to compounds that were not carcinogens in both mice and rats. This is perceived as the widest possible separation of the breast carcinogen-noncarcinogen classes.

Response to Technical Issue Raised in Year 1 Review

In the first Annual Review we indicated the creation of searchable databases for NTP and CPDB data. These databases consist of compiling the data from the CPDB and NTP into Excel spreadsheets that are easily viewed in order to identify compounds of particular interest (i.e., female-specific carcinogens). Although easily viewable, they have proven to be tedious and problematic for the creation of learning sets. To remedy this, we have developed a SAS routine for searching the data in its original format. We now have a more complete tool for learning set creation (SAS searching and Excel-based viewing of the data).

We also plan to shortly implement a routine to link the search results from SAS to a library of chemical structures for CPDB and NTP compounds. This will allow us to rapidly search for compounds with particular interest and instantly create files containing their chemical structures for SAR analysis in Sybyl. We think this will be of sufficient utility that we plan offer it to the NTP and CPDB or minimally provide it on our website.

Additionally, an added benefit to migrating from MCASE to Sybyl is that Sybyl allows the user to produce learning sets (i.e., databases of chemicals, names, and toxicological data) that are easily transportable between users and other systems. We therefore also plan to make publicly available on our website each of the specific model learning sets (e.g., mammary and female carcinogens learning sets.

Key Research Accomplishments
Development of female and mammary carcinogen models in MCASE
Development of mouse and rat mammary carcinogen models in Sybyl HQSAR-SIMCA
Ascertainment that Sybyl is an adequate replacement for MCASE

Reportable Outcomes
Seminar "Structure-activity relationships: Estrogen mimics and endocrine disruptors" at LSU Environmental Lecture Series at Tulane University

Conclusions
With the success of the first mammary carcinogen models we anticipate publishing these results in the near future (a proposed deliverable). If successful at modeling the general female carcinogens, another manuscript will be produced describing female-specific carcinogens (another deliverable).

To date, after technically about one year of work we have developed the proposed models set forth in Specific Aim 1 using MCASE. I estimate that we may be about one month behind schedule due to switching to Sybyl HQSAR-SIMCA. Moreover, in conjunction with this and other projects in my laboratory, all the required components for Specific Aim 2 are being moved from MCASE to Sybyl so there should be no delay or problems accomplishing the tasks of Specific Aim 2. This is of particular relevance for Specific Aims 2a and 2b that require other relevant (e.g., mutagenicity and estrogenicity) toxicological models on which to compare the female and mammary gland carcinogen models.

Looking forward, I see no obstacles to the successful completion of this project in a timely manner. By switching to Sybyl, we envision being able to more accurately and thoroughly investigate the chemical structural attributes of breast carcinogens.

References

1. Rosenkranz HS, Klopman G. The structural basis of the mutagenicity of chemicals in Salmonella typhimurium: The National Toxicology Program data base. Mutat Res 228:1-50(1990).

2. Zeiger E, Ashby J, Bakale G, Enslein K, Klopman G, Rosenkranz HS. Prediction of *Salmonella* mutagenicity. Mutagenesis 11:471-484(1996).

3. Cunningham AR, Cunningham SL, Rosenkranz HR. Structure Activity Approach to the Identification of Environmental Estrogens: The MCASE Approach. SAR QSAR Environ Res.